

What is Claimed is:

1 1. A computer-implemented method for randomly walking through
2 a hypertext-linked document set comprising a plurality of documents,
3 wherein at least a subset of the documents contain a plurality of links to
4 other documents, each document being associated with a host, the method
5 comprising:

6 a) selecting a host;
7 b) selecting at random a document associated with the host;
8 c) retrieving the selected document;
9 d) selecting at random a link in the retrieved document;
10 e) retrieving a document referenced by the selected link; and
11 f) repeating d) and e) until a predetermined condition is met.

1 2. The method of claim 1, further comprising, prior to d):

2 c.1) responsive to a random event:
3 c.1.1) selecting at random a host from among the previ-
4 ously selected hosts; and
5 c.1.2) repeating b) through f);

6 and wherein f) comprises repeating c.1) through e) until a predeter-
7 mined condition is met

1 3. The method of claim 1, further comprising, prior to d):

2 c.1) generating a random number;

3 c.2) determining whether the random number falls within a
4 predetermined range; and
5 c.3) responsive to the random number falling within the prede-
6 termined range:
7 c.1.1) selecting at random a host from among the previ-
8 ously selected hosts; and
9 c.1.2) repeating b) through f).

1 4. The method of claim 1, wherein the document set is the World
2 Wide Web, and wherein each document is a web page.

1 5. The method of claim 4, wherein each host corresponds to a do-
2 main.

1 6. The method of claim 1, further comprising, concurrently with a)
2 through f), performing a second two-level random walk through the hyper-
3 text-linked document set.

1 7. A computer-implemented method for randomly walking through
2 a hypertext-linked document set comprising a plurality of documents,
3 wherein at least a subset of the documents contain a plurality of links to
4 other documents, each document being associated with a host, the method
5 comprising:

6 a) initializing a host set;
7 b) initializing a document set for each host in the host set;
8 c) selecting at random a host from the host set;

9 d) selecting at random a document from the document set of
10 the selected host;
11 e) adding the selected host to the host set;
12 f) adding the selected document to the document set of the se-
13 lected host;
14 g) responsive to the selected document containing at least one
15 link:
16 g.1) selecting at random a link from the selected doc-
17 ument;
18 g.2) selecting a document corresponding to the selected
19 link;
20 g.3) selecting a host corresponding to the selected doc-
21 ument;
22 g.4) repeating e) through h) until a predetermined
23 condition is met; and
24 h) responsive to the selected document not containing at least
25 one link, repeating c) through h) until a predetermined con-
26 dition is met.

1 8. The method of claim 7, wherein:

2 e) is performed responsive to the selected host not being in the host
3 set; and
4 f) is performed responsive to the selected document not being in the
5 document set of the selected host.

1 9. The method of claim 7, wherein g) further comprises, prior to g.1):

603063-13742-100

2 g.0) responsive to a random event, repeating c) through h) until
3 a predetermined condition is met;

4 and wherein g.1) through g.4) are performed responsive to non-occur-
5 rence of the random event of g.0).

1 10. The method of claim 7, further comprising, prior to g.1):

2 g.0.1) generating a random number;

3 g.0.2) determining whether the random number falls within a
4 predetermined range; and

5 g.0.3) responsive to the random number falling within the prede-
6 termined range, repeating c) through h) until a predeter-
7 mined condition is met;

8 and wherein g.1) through g.4) are performed responsive to the ran-
9 dom number not falling within a predetermined range.

1 11. The method of claim 7, wherein the hypertext-linked document
2 set is the World Wide Web, and wherein each document is a web page.

1 12. The method of claim 11, wherein each host corresponds to a do-
2 main.

1 13. A computer-implemented method for measuring relative quality
2 of a search engine index, comprising:

3 a) performing a two-level random walk among documents
4 within a document set;

5 b) for each document encountered in the random walk, deter-
6 mining whether the document is indexed by the search en-
7 gine index; and
8 c) aggregating the results of b).

1 14. The method of claim 13, wherein at least a subset of the docu-
2 ments contain a plurality of links to other documents, each document being
3 associated with a host, and wherein a) comprises:

4 a.1) selecting a host;
5 a.2) selecting at random a document associated with the host;
6 a.3) retrieving the selected document;
7 a.4) selecting at random a link in the retrieved document;
8 a.5) retrieving a document referenced by the selected link; and
9 a.6) repeating a.4) and a.5) until a predetermined condition is
10 met.

1 15. The method of claim 14, further comprising, prior to a.4):

2 a.3.1) responsive to a random event:
3 a.3.1.1) selecting at random a host from among the previ-
4 ously selected hosts; and
5 a.3.1.2) repeating a.2) through a.6).

1 16. The method of claim 13, wherein at least a subset of the docu-
2 ments contain a plurality of links to other documents, each document being
3 associated with a host, and wherein a) comprises:

4 a.1) initializing a host set;

16. A method for selecting a document from a host set, comprising:

5 a.2) initializing a document set for each host in the host set;

6 a.3) selecting at random a host from the host set;

7 a.4) selecting at random a document from the document set of
the selected host;

8 a.5) adding the selected host to the host set;

9 a.6) adding the selected document to the document set of the se-
lected host;

10 a.7) responsive to the selected document containing at least one
link:

11 a.7.1) selecting at random a link from the selected doc-
ument;

12 a.7.2) selecting a document corresponding to the selected
link;

13 a.7.3) selecting a host corresponding to the selected doc-
ument;

14 a.7.4) repeating a.5) through a.8) until a predetermined
condition is met; and

15 a.8) responsive to the selected document not containing at least
one link, repeating a.3) through a.8) until a predetermined
condition is met.

1 17. The method of claim 16, wherein:

2 a.5) is performed responsive to the selected host not being in the host
set; and

4 a.6) is performed responsive to the selected document not being in the
document set of the selected host.

1 18. The method of claim 13, wherein each document contains a plu-
2 rality of words, and wherein b) comprises, for each document encountered in
3 the random walk:

4 b.1) selecting at least one word from the document;
5 b.2) performing a query on the search engine index based on the
6 selected at least one word, to obtain search results; and
7 b.3) determining whether the document is included in the ob-
8 tained search results.

1 19. The method of claim 18, wherein b.1) comprises selecting at least
2 one word based on rarity.

1 20. A computer-implemented method for measuring relative quality
2 of a document in a document set, comprising:

3 a) performing a two-level random walk among documents
4 within a document set; and
5 b) determining a quality metric responsive to the number of
6 times the document is encountered in the random walk.

1 21. A computer-implemented method for measuring relative quality
2 of a document in a document set comprising a plurality of documents,
3 wherein at least a subset of the documents contain a plurality of links to
4 other documents, the method comprising:

5 a) performing a two-level random walk among documents
6 within a document set; and

7 b) determining a quality metric responsive to the number of
8 documents that link to the document.

1 22. The method of claim 21, wherein b) comprises determining a qual-
2 ity metric responsive to the number of documents that link to the docu-
3 ment, and responsive to the quality metric of the linking documents.

1 23. The method of claim 21, wherein b) comprises determining a
2 value for:

3
$$R(p) = d / T + (1 - d) \sum_{i=1}^k R(p_i) / C(p_i)$$

4 where:

5 T is the total number of documents in the document set;
6 d is a damping factor such that $0 < d < 1$;
7 documents p_1, \dots, p_k each contain at least one link to document p; and
8 C(p) is the number of links out of p.

1 24. The method of claim 21, wherein each document is associated
2 with a host, and wherein a) comprises:

- 3 a.1) selecting a host;
- 4 a.2) selecting at random a document associated with the host;
- 5 a.3) retrieving the selected document;
- 6 a.4) responsive to a random event:
 - 7 a.4.1) selecting at random a host from among the previ-
 - 8 ously selected hosts; and
 - 9 a.4.2) repeating a.2) through a.7);
- 10 a.5) selecting at random a link in the retrieved document;

11 a.6) retrieving a document referenced by the selected link; and
12 a.7) repeating a.4) to a.6) until a predetermined condition is met.

1 25. The method of claim 21, wherein each document is associated
2 with a host, and wherein a) comprises:

3 a.1) initializing a host set;

4 a.2) initializing a document set for each host in the host set;

5 a.3) selecting at random a host from the host set;

6 a.4) responsive to a random event:
7 a.4.1) selecting at random a host from among the previ-
8 ously selected hosts; and
9 a.4.2) repeating a.2) through a.7).

10 a.5) selecting at random a document from the document set of
11 the selected host;

12 a.6) adding the selected host to the host set;

13 a.7) adding the selected document to the document set of the se-
14 lected host;

15 a.8) responsive to the selected document containing at least one
16 link:
17 a.8.1) selecting at random a link from the selected doc-
18 ument;
19 a.8.2) selecting a document corresponding to the selected
20 link;
21 a.8.3) selecting a host corresponding to the selected doc-
22 ument; and

23 a.8.4) repeating a.6) through a.9) until a predetermined
24 condition is met; and

25 a.9) responsive to the selected document not containing at least
26 one link, repeating a.3) through a.9) until a predetermined
27 condition is met.

1 26. The method of claim 21, further comprising:

2 c) determining a quality metric for at least one additional doc-
3 ument; and
4 d) ranking the quality metric of the first document with respect
5 to the quality metrics of the additional documents.

1 27. A computer-implemented method for randomly walking through
2 a hypertext-linked document set comprising a plurality of documents,
3 wherein at least a subset of the documents contain a plurality of links to
4 other documents, each document being associated with a host, the method
5 comprising:

6 a) selecting a host;
7 b) selecting at random a document associated with the host;
8 c) retrieving the selected document;
9 d) responsive to a random event:
10 d.1) selecting at random a host from among the previ-
11 ously selected hosts; and
12 d.2) repeating b) through e) until a predetermined con-
13 dition is met
14 e) responsive to the random event not occurring:

15 e.1) selecting at random a link in the retrieved docu-
16 ment;
17 e.2) retrieving a document referenced by the selected
18 link; and
19 e.3) repeating d) and e) until a predetermined condi-
20 tion is met.

1 28. A computer-implemented method for measuring relative quality
2 of a document in a document set comprising a plurality of documents,
3 wherein at least a subset of the documents contain a plurality of links to
4 other documents, the method comprising:

5 a) performing a two-level random walk among documents
6 within a document set, the two-level random walk compris-
7 ing:
8 a.1) initializing a host set;
9 a.2) initializing a document set for each host in the host
10 set;
11 a.3) selecting at random a host from the host set;
12 a.4) responsive to a random event:
13 a.4.1) selecting at random a host from among the
14 previously selected hosts; and
15 a.4.2) repeating a.2) through a.7).
16 a.5) selecting at random a document from the document
17 set of the selected host;
18 a.6) adding the selected host to the host set;

卷之三

~~1 29. A computer program product comprising a computer-usab
2 le medium having computer-readable code embodied therein for randomly
3 walking through a hypertext-linked document set comprising a plurality of
4 documents, wherein at least a subset of the documents contain a plurality of~~

links to other documents, each document being associated with a host, the computer program product comprising:

- a) computer-readable program code devices configured to cause a computer to select a host;
- b) computer-readable program code devices configured to cause a computer to select at random a document associated with the host;
- c) computer-readable program code devices configured to cause a computer to retrieve the selected document;
- d) computer-readable program code devices configured to cause a computer to select at random a link in the retrieved document;
- e) computer-readable program code devices configured to cause a computer to retrieve a document referenced by the selected link; and
- f) computer-readable program code devices configured to cause a computer to repeat the operations of d) and e) until a pre-determined condition is met.

30. The computer program product of claim 29, further comprising computer-readable program code devices configured to cause a computer to, prior to selecting at random a link in the retrieved document:

- c.1) responsive to a random event:
 - select at random a host from among the previously selected hosts; and
 - repeat the operations of b) through f);

8 and wherein the computer-readable program code devices configured
9 to cause a computer to repeat the operations of d) and e) until a predeter-
10 mined condition is met comprise computer-readable program code devices
11 configured to cause a computer to repeat the operations of c.1) through e) un-
12 til a predetermined condition is met.

1 31. The computer program product of claim 29, further comprising:
2 computer-readable program code devices configured to cause a com-
3 puter to generate a random number;
4 computer-readable program code devices configured to cause a com-
5 puter to determine whether the random number falls
6 within a predetermined range; and
7 computer-readable program code devices configured to cause a com-
8 puter to, responsive to the random number falling within
9 the predetermined range:
10 select at random a host from among the previously selected
11 hosts; and
12 repeat the operations of b) through f).

1 32. The computer program product of claim 29, wherein the docu-
2 ment set is the World Wide Web, and wherein each document is a web page.

1 33. The computer program product of claim 32, wherein each host
2 corresponds to a domain.

1 34. The computer program product of claim 29, further comprising
2 computer-readable program code devices configured to cause a computer to,

3 concurrently with the operations of a) through f), perform a second two-
4 level random walk through the hypertext-linked document set.

1 35. A computer program product comprising a computer-usuable
2 medium having computer-readable code embodied therein for randomly
3 walking through a hypertext-linked document set comprising a plurality of
4 documents, wherein at least a subset of the documents contain a plurality of
5 links to other documents, each document being associated with a host, the
6 computer program product comprising:

- 7 a) computer-readable program code devices configured to cause
8 a computer to initialize a host set;
- 9 b) computer-readable program code devices configured to cause
10 a computer to initialize a document set for each host in the
11 host set;
- 12 c) computer-readable program code devices configured to cause
13 a computer to select at random a host from the host set;
- 14 d) computer-readable program code devices configured to cause
15 a computer to select at random a document from the docu-
16 ment set of the selected host;
- 17 e) computer-readable program code devices configured to cause
18 a computer to add the selected host to the host set;
- 19 f) computer-readable program code devices configured to cause
20 a computer to add the selected document to the document
21 set of the selected host;

Case 3792

22 g) computer-readable program code devices configured to cause
23 a computer to, responsive to the selected document contain-
24 ing at least one link:
25 g.1) select at random a link from the selected docu-
26 ment;
27 g.2) select a document corresponding to the selected
28 link;
29 g.3) select a host corresponding to the selected docu-
30 ment; and
31 g.4) repeat the operations of e) through h) until a pre-
32 determined condition is met; and
33 h) computer-readable program code devices configured to cause
34 a computer to, responsive to the selected document not con-
35 taining at least one link, repeat the operations of c) through
36 h) until a predetermined condition is met.

1 36. The computer program product of claim 35, wherein:
2 the computer-readable program code devices configured to cause a
3 computer to add the selected host to the host set operate re-
4 sponsive to the selected host not being in the host set; and
5 the computer-readable program code devices configured to cause a
6 computer to add the selected document to the document set
7 of the selected host operate responsive to the selected docu-
8 ment not being in the document set of the selected host.

1 37. The computer program product of claim 35, wherein computer-
2 readable program code devices g) further comprise computer-readable pro-
3 gram code devices configured to cause a computer to, prior to g.1):

4 g.0) responsive to a random event, repeat the operations of c)
5 through h) until a predetermined condition is met;

6 and wherein computer-readable program code devices g) are config-
7 ured to cause a computer to perform g.1) through g.4) responsive to non-oc-
8 currence of the random event of g.0).

1 38. The computer program product of claim 35, wherein computer-
2 readable program code devices g) further comprise computer-readable pro-
3 gram code devices configured to cause a computer to, prior to g.1):

4 g.0.1) generate a random number;

5 g.0.2) determine whether the random number falls within a pre-
6 determined range; and

7 g.0.3) responsive to the random number falling within the pre-
8 determined range, repeat the operations of c) through h) until
9 a predetermined condition is met;

10 and wherein computer-readable program code devices g) are config-
11 ured to cause a computer to perform g.1) through g.4) responsive to the ran-
12 dom number not falling within a predetermined range.

1 39. The computer program product of claim 35, wherein the hyper-
2 text-linked document set is the World Wide Web, and wherein each docu-
3 ment is a web page.

1 40. The computer program product of claim 39, wherein each host
2 corresponds to a domain.

1 41. A computer program product comprising a computer-readable
2 medium having computer-readable code embodied therein for measuring
3 relative quality of a search engine index, the computer program product
4 comprising:

- 5 a) computer-readable program code devices configured to cause
6 a computer to perform a two-level random walk among
7 documents within a document set;
- 8 b) computer-readable program code devices configured to cause
9 a computer to, for each document encountered in the ran-
10 dom walk, determine whether the document is indexed by
11 the search engine index; and
- 12 c) computer-readable program code devices configured to cause
13 a computer to aggregate the results of the operations of b).

1 42. The computer program product of claim 41, wherein at least a sub-
2 set of the documents contain a plurality of links to other documents, each
3 document being associated with a host, and wherein the computer-readable
4 program code devices configured to cause a computer to perform a two-level
5 random walk comprise:

卷一百一十五

1 43. The computer program product of claim 42, further comprising
2 computer-readable program code devices configured to cause a computer to,
3 prior to selecting at random a link in the retrieved document:

4 a.3.1) responsive to a random event:
5 select at random a host from among the previously selected
6 hosts; and
7 repeat the operations of a.2) through a.6).

1 44. The computer program product of claim 41, wherein at least a sub-
2 set of the documents contain a plurality of links to other documents, each
3 document being associated with a host, and wherein the computer-readable
4 program code devices configured to cause a computer to perform a two-level
5 random walk comprise:

6 a.1) computer-readable program code devices configured to cause
7 a computer to initialize a host set;

8 a.2) computer-readable program code devices configured to cause
9 a computer to initialize a document set for each host in the
10 host set;

11 a.3) computer-readable program code devices configured to cause
12 a computer to select at random a host from the host set;

13 a.4) computer-readable program code devices configured to cause
14 a computer to select at random a document from the docu-
15 ment set of the selected host;

16 a.5) computer-readable program code devices configured to cause
17 a computer to add the selected host to the host set;

18 a.6) computer-readable program code devices configured to cause
19 a computer to add the selected document to the document
20 set of the selected host;

21 a.7) computer-readable program code devices configured to cause
22 a computer to, responsive to the selected document contain-
23 ing at least one link:
24 a.7.1) select at random a link from the selected docu-
25 ment;

26 a.7.2) select a document corresponding to the selected
27 link;
28 a.7.3) select a host corresponding to the selected docu-
29 ment;
30 a.7.4) repeat the operations of a.5) through a.8) until a
31 predetermined condition is met; and
32 a.8) computer-readable program code devices configured to cause
33 a computer to, responsive to the selected document not con-
34 taining at least one link, repeat the operations of a.3)
35 through a.8) until a predetermined condition is met.

1 45. The computer program product of claim 44, wherein:
2 the computer-readable program code devices configured to cause a
3 computer to add the selected host to the host set are config-
4 ured to cause a computer to add the selected host responsive
5 to the selected host not being in the host set; and
6 the computer-readable program code devices configured to cause a
7 computer to add the selected document to the document set
8 of the selected host are configured to cause a computer to
9 add the selected document responsive to the selected docu-
10 ment not being in the document set of the selected host.

1 46. The computer program product of claim 41, wherein each docu-
2 ment contains a plurality of words, and wherein the computer-readable pro-
3 gram code devices configured to cause a computer to, determine whether the
4 document is indexed by the search engine index comprise computer-readable

5 program code devices configured to, for each document encountered in the
6 random walk:

7 b.1) select at least one word from the document;
8 b.2) perform a query on the search engine index based on the se-
9 lected at least one word, to obtain search results; and
10 b.3) determine whether the document is included in the ob-
11 tained search results.

1 47. The computer program product of claim 46, wherein the com-
2 puter-readable program code devices configured to select at least one word
3 from the document comprise computer-readable program code devices con-
4 figured to select at least one word based on rarity.

1 48. A computer program product comprising a computer-readable
2 medium having computer-readable code embodied therein for measuring
3 relative quality of a document in a document set, the computer program
4 product comprising:

5 computer-readable program code devices configured to cause a com-
6 puter to perform a two-level random walk among docu-
7 ments within a document set; and
8 computer-readable program code devices configured to cause a com-
9 puter to determine a quality metric responsive to the num-
10 ber of times the document is encountered in the random
11 walk.

Case 3792

1 49. A computer program product comprising a computerUsable
2 medium having computer-readable code embodied therein for measuring
3 relative quality of a document in a document set comprising a plurality of
4 documents, wherein at least a subset of the documents contain a plurality of
5 links to other documents, the computer program product comprising:

6 computer-readable program code devices configured to cause a com-
7 puter to perform a two-level random walk among docu-
8 ments within a document set; and

9 computer-readable program code devices configured to cause a com-
10 puter to determine a quality metric responsive to the num-
11 ber of documents that link to the document.

1 50. The computer program product of claim 49, wherein the com-
2 puter-readable program code devices configured to cause a computer to de-
3 termine a quality metric comprise computer-readable program code devices
4 configured to cause a computer to determine a quality metric responsive to
5 the number of documents that link to the document, and responsive to the
6 quality metric of the linking documents.

1 51. The computer program product of claim 49, wherein the com-
2 puter-readable program code devices configured to cause a computer to de-
3 termine a quality metric comprise computer-readable program code devices
4 configured to cause a computer to determine a value for:

$$5 \quad R(p) = d / T + (1 - d) \sum_{i=1}^k R(p_i) / C(p_i)$$

6 where:

7 T is the total number of documents in the document set;
8 d is a damping factor such that $0 < d < 1$;
9 documents p_1, \dots, p_k each contain at least one link to document p; and
10 $C(p)$ is the number of links out of p.

1 52. The computer program product of claim 49, wherein each docu-
2 ment is associated with a host, and wherein the computer-readable program
3 code devices configured to cause a computer to perform a two-level random
4 walk comprise:

- 5 a.1) computer-readable program code devices configured to cause
6 a computer to select a host;
- 7 a.2) computer-readable program code devices configured to cause
8 a computer to select at random a document associated with
9 the host;
- 10 a.3) computer-readable program code devices configured to cause
11 a computer to retrieve the selected document;
- 12 a.4) computer-readable program code devices configured to cause
13 a computer to, responsive to a random event:
 - 14 a.4.1) select at random a host from among the previ-
15 ously selected hosts; and
 - 16 a.4.2) repeat the operations of a.2) through a.7);
- 17 a.5) computer-readable program code devices configured to cause
18 a computer to select at random a link in the retrieved doc-
19 ument;

20 a.6) computer-readable program code devices configured to cause
21 a computer to retrieve a document referenced by the selected
22 link; and
23 a.7) computer-readable program code devices configured to cause
24 a computer to repeat the operations of a.4) to a.6) until a pre-
25 determined condition is met.

1 53. The computer program product of claim 49, wherein each docu-
2 ment is associated with a host, and wherein and wherein the computer-
3 readable program code devices configured to cause a computer to perform a
4 two-level random walk comprise:

5 a.1) computer-readable program code devices configured to cause
6 a computer to initialize a host set;
7 a.2) computer-readable program code devices configured to cause
8 a computer to initialize a document set for each host in the
9 host set;
10 a.3) computer-readable program code devices configured to cause
11 a computer to select at random a host from the host set;
12 a.4) computer-readable program code devices configured to cause
13 a computer to, responsive to a random event:
14 a.4.1) select at random a host from among the previ-
15 ously selected hosts; and
16 a.4.2) repeat the operations of a.2) through a.7).
17 a.5) computer-readable program code devices configured to cause
18 a computer to select at random a document from the docu-
19 ment set of the selected host;

SEARCHED

20 a.6) computer-readable program code devices configured to cause
21 a computer to add the selected host to the host set;
22 a.7) computer-readable program code devices configured to cause
23 a computer to add the selected document to the document
24 set of the selected host;
25 a.8) computer-readable program code devices configured to cause
26 a computer to, responsive to the selected document contain-
27 ing at least one link:
28 a.8.1) select at random a link from the selected docu-
29 ment;
30 a.8.2) select a document corresponding to the selected
31 link;
32 a.8.3) select a host corresponding to the selected docu-
33 ment; and
34 a.8.4) repeat the operations of a.6) through a.9) until a
35 predetermined condition is met; and
36 a.9) responsive to the selected document not containing at least
37 one link, repeating the operations of a.3) through a.9) until a
38 predetermined condition is met.

1 54. The computer program product of claim 49, further comprising:
2 c) computer-readable program code devices configured to cause
3 a computer to determine a quality metric for at least one ad-
4 ditional document; and
5 d) computer-readable program code devices configured to cause
6 a computer to rank the quality metric of the first document

7 with respect to the quality metrics of the additional docu-
8 ments.

1 55. A computer program product comprising a computer-usuable
2 medium having computer-readable code embodied therein for randomly
3 walking through a hypertext-linked document set comprising a plurality of
4 documents, wherein at least a subset of the documents contain a plurality of
5 links to other documents, each document being associated with a host, the
6 computer program product comprising:

- 7 a) computer-readable program code devices configured to cause
8 a computer to select a host;
- 9 b) computer-readable program code devices configured to cause
10 a computer to select at random a document associated with
11 the host;
- 12 c) computer-readable program code devices configured to cause
13 a computer to retrieve the selected document;
- 14 d) computer-readable program code devices configured to cause
15 a computer to, responsive to a random event:
 - 16 d.1) select at random a host from among the previ-
17 ously selected hosts; and
 - 18 d.2) repeat the operations of b) through e) until a pre-
19 determined condition is met
- 20 e) computer-readable program code devices configured to cause
21 a computer to, responsive to the random event not occur-
22 ring:
 - 23 e.1) select at random a link in the retrieved document;

- e.2) retrieve a document referenced by the selected link; and
- e.3) repeat the operations of d) and e) until a predetermined condition is met.

56. A computer program product comprising a computer-usuable medium having computer-readable code embodied therein for measuring relative quality of a document in a document set comprising a plurality of documents, wherein at least a subset of the documents contain a plurality of links to other documents, the computer program product comprising:

a) computer-readable program code devices configured to cause a computer to perform a two-level random walk among documents within a document set, the computer-readable program code devices comprising:

a.1) computer-readable program code devices configured to cause a computer to initialize a host set;

a.2) computer-readable program code devices configured to cause a computer to initialize a document set for each host in the host set;

a.3) computer-readable program code devices configured to cause a computer to select at random a host from the host set;

a.4) computer-readable program code devices configured to cause a computer to, responsive to a random event:

a.4.1) select at random a host from among the previously selected hosts; and

22 a.4.2) repeat the operations of a.2) through a.7).

23 a.5) computer-readable program code devices configured

24 to cause a computer to select at random a document

25 from the document set of the selected host;

26 a.6) computer-readable program code devices configured

27 to cause a computer to add the selected host to the

28 host set;

29 a.7) computer-readable program code devices configured

30 to cause a computer to add the selected document to

31 the document set of the selected host;

32 a.8) computer-readable program code devices configured

33 to cause a computer to, responsive to the selected

34 document containing at least one link:

35 a.8.1) select at random a link from the selected

36 document;

37 a.8.2) select a document corresponding to the se-

38 lected link;

39 a.8.3) select a host corresponding to the selected

40 document;

41 a.8.4) repeat the operations of a.6) through a.9) un-

42 til a predetermined condition is met; and

43 a.9) computer-readable program code devices configured

44 to cause a computer to, responsive to the selected

45 document not containing at least one link, repeat the

46 operations of a.3) through a.9) until a predetermined

47 condition is met;

48 b) computer-readable program code devices configured to cause
49 a computer to determine a quality metric responsive to the
50 number of documents that link to the document;
51 c) computer-readable program code devices configured to cause
52 a computer to determine a quality metric for at least one ad-
53 ditional document; and
54 d) computer-readable program code devices configured to cause
55 a computer to rank the quality metric of the first document
56 with respect to the quality metrics of the additional docu-
57 ments.

1 57. A system for randomly walking through a hypertext-linked doc-
2 ument set comprising a plurality of documents, wherein at least a subset of
3 the documents contain a plurality of links to other documents, each docu-
4 ment being associated with a host, the system comprising:

5 a) a host selector;
6 b) a random document selector, coupled to the host selector,
7 for selecting at random a document associated with the host;
8 c) a document retriever, coupled to the random document se-
9 lector, for retrieving the selected document; and
10 d) a link selector, coupled to the document retriever, for select-
11 ing at random a link in the retrieved document;

12 wherein the document retriever retrieves a document referenced by
13 the selected link;

14 and wherein the link selector repeatedly selects at random a link and
15 the document retriever repeatedly retrieves a document referenced by the se-
16 lected link, until a predetermined condition is met.

1 58. A system for measuring relative quality of a search engine index,
2 comprising:

3 a random walker, for performing a two-level random walk among
4 documents within a document set;
5 a determination module, coupled to the random walker, for, for each
6 document encountered in the random walk, determining
7 whether the document is indexed by the search engine in-
8 dex; and

9 a results aggregation module, coupled to the determination module,
10 for aggregating the results of the determination module.

1 59. A system for measuring relative quality of a document in a docu-
2 ment set, comprising:

3 a random walker, for performing a two-level random walk among
4 documents within a document set; and
5 a determination module, coupled to the random walker, for deter-
6 mining a quality metric responsive to the number of times
7 the document is encountered in the random walk.

A handwritten signature consisting of stylized initials and a surname, written in black ink.